DOCUMENT RESUME

ED 053 208 TM 000 741

AUTHOR Theimer, William C., Jr.
TITLE Alternatives to the State Mandated Testing Program.
PUB DATE Apr 71
NOTE 22p.; From part of a symposium on Testing at the Convention of the California Educational Research Association, San Diego, California, April 29, 1971

EDRS PRICE EDRS Price MF-$0.65 HC-$3.29
DESCRIPTORS *Criterion Referenced Tests, Discriminant Analysis, Educational Innovation, Feedback, *Norm Referenced Tests, Norms, Objectives, Parents, Program Evaluation, Sampling, School Personnel, Standardized Tests, *State Programs, *Testing Programs, *Test Interpretation, Tests
IDENTIFIERS Instructional Objective Exchange

ABSTRACT

The purposes for which test data should be used were described as: test should provide a learning experience for the child taking the test, the teacher giving the test, and the parents of the child. It should also provide information useful to school administrators, school boards, and state and national educational agencies. The state-mandated testing program was examined in terms of how well it provided information which would be most valuable to these groups. Students, teachers and parents get very little helpful information from the present testing program. The information provided to the last three groups mentioned was more useful, but wasteful in time and money. Three alternatives were suggested: (1) criterion referenced tests based on the UCLA Instructional Objectives Exchange (IOX); (2) a form of sampling using the present standardized state-mandated tests; and (3) a combination of standardized normative based tests and locally constructed criterion referenced tests. (Author/AG)

ED053208

Abstract

Alternatives to the State Mandated Testing Program

The purposes for which test data should be used were described as: test should provide a learning experience for the child taking the test, the teacher giving the test, and the parents of the child. It should also provide information useful to school administrators, school boards, and state and national educational agencies.

The state-mandated testing program was examined in terms of how well it provided information which would be most valuable to these groups. Students, teachers and parents get very little helpful information from the present testing program. The information provided to the last three groups mentioned was more useful, but wasteful in time and money.

Three alternatives were mentioned. First, criterion referenced tests based on the UCLA Instructional Objectives Exchange (IOX) were suggested as being more appropriate for students, teachers and parents to get information which was relevant to the individual child's progress. Second, a form of sampling using the present standardized state-mandated tests was suggested to avoid the invidious comparisons now made between schools and school districts. Third, a combination of standardized normative based tests and locally constructed criterion referenced tests was suggested as a way of providing the maximum information to all six groups mentioned in paragraph one.

TM 000 241

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY

Alternatives to the State Mandated Testing Program

An address given as part of a Symposium on Testing at the CERA Convention, April 29, 1971

Wm. C. Theimer, Jr.
University of the Pacific

The purpose of any testing program is to provide information. As Dr. Castenada and Dr. Nelson have indicated, the information provided by the normative-based tests required by the State of California do not provide valid information about certain groups of individuals in particular. I would like to examine the problems which the program poses in providing useful information to others who are required to administer the testing program.

First, since the purpose of a testing program is to provide information to a wide variety of people (Lyman, 1963, Dyer, 1970) let us examine just how useful the information provided by the state-mandated program is to each of these users. Tests should be used as a teaching-learning technique (Storey, 1970). This statement implies and Dyer states that the two most important groups for whom the tests should be providing information are the pupils who take the tests and the teachers of these children. How much information is really given to the student by the state-mandated tests? In most instances, very little. Norm-referenced tests do not provide the kind of feedback necessary to permit a student to know the specific skills or knowledge which he is lacking, and more importantly, how much he has learned (Millman, 1971, Tyler, 1970). This lack is particularly acute with the disadvantaged child, as Tyler has so adequately

pointed out, and yet it is this group which is most often tested with these normative based tests. The feedback is usually provided in grade placement scores which are easily misinterpretable and in many instances invalid (McAloon, 1969) and unreliable (Stake and Wardrop, 1971).

Since the grade placement score is so widely misinterpreted, and since the items which are included in the normative based tests are not an adequate sample of the kinds of learnings which can be expected of disadvantaged students, the tests provide very little in the way of helpful information for these students. In fact, if incorrectly used, as they frequently are, they can be harmful (McAloon, 1970, Glasser, 1969).

Since the items which are used to make a good normative based test are chosen to discriminate between the high achieving and low achieving child, they perform a ranking function. That is, the discrimination index which is most important to achieve for each item is between .40 and .60 (Storey, 1970, Lyman, 1963). When a child is compared to his peers in terms of some norm, usually the mean of the reference group, those in the bottom 50% can easily feel that they are "failures".

These two factors, ranking and feelings of failure, can have a deleterious effect on a large per cent of the students since at least half of them will fall below the mean as a simple artifact of the test. The failure syndrome which is prevalent in our most disadvantaged schools and which we try to teach our young pre-service teacher

trainees to avoid is reinforced by the administration of normative based tests which frequently do not test the objectives the teacher has determined are most realistic in terms of the needs of the individual child. Another problem with these tests is that, of necessity, they use only a small sample of all the possible learnings students at a particular level have made during the year. Thus, the disadvantaged or slow pupil is able to demonstrate only a limited amount of the skills and knowledge he has gained. There are many other areas which could be cited, such as regression, S.E. misinterpretation, etc., to indicate that this kind of testing is inappropriate for providing maximum information to pupils. The preceding will suffice, however, to indicate that the state-mandated tests do a very poor job of providing useful information to one of the most important groups of people who should be receiving this information--the pupils.

A second important person for whom information about the pupil is necessary is the teacher. Here, the state mandated tests also rate poorly, although some school districts such as the Stockton Unified School District are making maximum use of the results by providing information about the pupils in her class to each teacher. By using stanines and comparing achievement test data with measures of ability, the teacher can be alerted to certain problems which should be examined more carefully. Unfortunately, the S.E. of measurement and the regression effect would predict that on

normative based tests some children would appear in the extremes on one or the other of these tests by chance, and that some of the children whose scores appear in the "appropriate" stanine groups will have done so because of errors of measurement. This use of the tests does alert the teacher to possible problems, however, and can serve a useful function.

An individual profile is also provided to teachers where discrepancies between expected and actual achievement scores differ significantly for the child. Here again, the teacher does gain some knowledge about possible problem children. Unfortunately, since the normative based test must, of necessity, sample only a small portion of each of the major areas being taught, the teacher must resort to other tests of a diagnostic nature to determine just where these children are failing and where they are accomplishing. Other alternatives which will be mentioned later would preclude her having to perform this additional effort.

A third group for whom information should be provided is the school administrators and local school boards. Here, the information from the state-mandated testing program does somewhat better than in the two previous categories. The normative-based test provides information which is more in the nature of a ranking than in terms of precisely what has been learned. Two children achieving the same grade placement score could have answered correctly items that were quite different. That is, the specific items each answered correctly might not overlap at all.

These tests do not tell precisely what a group has learned. (McAloon, 1970).

The school board can, however, determine from this information how the various schools within the district are progressing by comparing the mean scores of schools within the district from year to year. They can also see how the schools within their district compare with some normative group, usually referred to as "the national average". Unfortunately, much information which should also be provided along with these data to make them meaningful such as beginning level of achievement, socio-economic data, teacher proficiency, etc. (Barro, 1971, Dyer, 1971) is omitted, and the comparisons are used incorrectly. This failure to interpret scores properly causes school administrators, teachers, and parents to become indignant when the scores of individual schools are printed in the local press and they frequently accuse each other of causing the failure of the children to learn. Much fire is generated, but little light is provided which will help improve the instructional program. When innovative ideas are being tried, the actual amount that was learned is not reflected in these tests since the items in the test are not valid--that is, they are not testing the program which is actually being carried out in the schools. Schools which are individualizing instruction are at a particular disadvantage when a single test is used to measure all schools without any consideration of the imputs or constraints placed on the various schools (Barro, 1971).

Unfortunately, most school boards expect that their

schools must meet or exceed the norms of the reference group without taking into consideration the composition of the reference group. The fact that nationally normed tests are not comparable in terms of the standard scores reported, usually the grade placement score, has been pointed out in earlier research (Taylor and Crandall, 1962) conducted in California. By simply changing from one nationally standardized test to another, schools and school districts can suddenly "gain" in the amount the students have learned. As we all know, this is partially a result of the group on whom the test was normed and partially by the choice of items within the different forms of the test.

Thus, although the state-mandated tests do provide school boards and school administrators some guidelines as to the relative standing of their schools in comparison with other groups, this information taken alone is frequently used inappropriately and causes much unnecessary dissension and fear among the teaching ranks and criticism of the schools by parents. The information provided these groups, then, by this testing program is a mixed blessing.

A fourth group who are vitally interested in test data are the parents. We have, unfortunately, in the past conditioned them to accept the A, B, C grading system as providing them with good information about what their children are learning. They have been trained to look for comparative data rather than the actual achievement of their children in terms of specific skills and knowledges learned. The information provided them from the state testing program

is frequently given in either grade-placement scores or percentile ranks without any prior explanation as to what these mean. A parent of a fifth grade student who is told that his child is reading at the fourth grade level is rightly indignant and frequently critical of the schools. He may know that his child is reading some things now that he was not reading at the beginning of the year, but he forgets this in his fear that his child may be a hopeless failure all his life. The fact that the 4.0 grade placement score is an extrapolation is never made clear to him, and the fact that his child has progressed well during the year in learning to read by increasing his vocabulaty, or learning to read by some phonics approach, is not indicated by this single number.

While I think most of us would agree that it is important that parents know the relative standing of their children so that they will not have unreal expectations of their future progress through our academically oriented public schools, the fact that this information is given in grades one and two may have a more harmful effect than any good that it does. Piaget and other psychologists and educators who think in terms of developmental processes in the child have frequently pointed out that the normal curve indicates that maturation in any particular skill or phase of development is not accomplished by all children at the same time. More will be said regarding this point later in the paper, when alternatives are mentioned. The problems which I have enumerated above are sufficient to indicate that the state-mandated, normative referenced tests

do not provide parents with the most useful information that they need. These tests provide data which compares their child with other children rather than concentrating on the amount and quality of learning accomplished. Better methods of providing this information are available.

A last group for which the test information is useful is the State Board of Education and the State legislature. Here the normative based tests provide the most useful information since the test data is being used in the manner for which it is most appropriate. That is, the State Board of Education and the State legislature are interested in how well the schools are progressing in terms of some general goals or objectives which are set by the state. The data from these normative based tests, when compared from year to year and across districts, does provide them with useful information in terms of how well the children in the state are progressing, provided that the tests used are valid for the groups being tested. There is serious question as to the validity of the state mandated tests for large segments of our population who are presently included in the testing program. Others, such as those groups Tyler has indicated, usually referred to as the Title I or disadvantaged students, should also be included in the groups for which these tests are frequently invalid, and for which only guessing scores are being compiled. The standardized test does provide useful information, even given these problems in providing bench-marks about how well the children are progressing in terms of the broad objectives when compared to other regions and to scores of students who were in school in previous years.

There are some problems even here, however. First, as mentioned earlier, a change from one normative based test to another will show an artificial gain or loss, depending on the test chosen, in the standing of the children across the state. These gains or losses are simply artifacts of the sample of items chosen to measure certain areas of the curriculum, such as reading or mathematical concepts, and of the sample of children used to develop the norms. A second problem is that changes in education which are being implemented in the schools are not reflected in the standardized test items. Although test makers are always incorporating new changes in curriculum in their tests, such as the "new math" items when SMSG came along or biology items for BSCS Biology, these changes are usually not reflected in the tests for two or three years following the introduction of this new concept in the most innovative schools. In addition, certain districts are much more willing to innovate and try the new than others. The state-mandated testing program might show these districts as doing a very poor job since the tests do not really measure the objectives of that district, when in fact the district is meeting its objectives very well. In other words, the objectives of the tests on which the normative data are based must be congruent with those of the local district or the tests are not validly measuring the education of the children within that district. The use of one single measure as the criteria by which schools are to be judged will tend to inhibit innovation and research which are sorely needed in our schools (Dyer, 1970).

Given that the state-mandated tests do provide useful information to the state officials, the testing of all children in the grades specified every year is unnecessary. By using a ten per cent sampling (Sax, 1968, p. 143) of all the students in the grades being tested as mandated by the state, the same kind of information could be obtained as is now obtained by testing all the children. The loss of time for teaching used in giving these tests, especially to groups for which they are invalid, plus the cost of the tests themselves makes this testing a great waste of both valuable teaching time and money.

ALTERNATIVES:

Since the state-mandated testing program does not provide the kind of information which is most useful to students, teachers, parents, and information which is only partially useful to local school administrators and school boards, what kind of alternatives are available which could provide better information to the state and these other users? First, I would like to examine the basic purpose of education in the United States. How does education here differ from that in the USSR, for example, for this basic purpose underlies the choice of objectives which are to be used in the selection of tests and how the test information is to be used. One of the areas is in terms of promoting as much diversity, individuality, and creativity as possible in our people. I would like to quote from Sterling McMurrin, former U.S. Commissioner of Education at this point:

> "It is a question whether in the future there will be a society in which the autonomous person is both the sovereign and sacred end, or a technological order in which individualism will yield to collectivism, freedom to regimentation, diversity to boredome of sameness, and where persons will be transformed into things." (1971)

Standardized tests tend to restrict the innovative in education, and teachers are frequently tempted to "teach to the test as a consequence (Mollenkamp, 1970). To avoid this, I would like to suggest that the Instructional Objectives Exchange (IOX) offers several possible alternatives which will more nearly meet the goals of education as it is usually espoused--the development of automomous individuals.

If we really mean what we say when we talk about individualizing instruction, then we must have some kind of measure which will not penalize the schools which are trying to develop new approaches to the teaching of children or conduct meaningful research in the schools. In both cases, the objectives which are stated by those who are doing the teaching should be what is measured by the test. Since many of these programs will offer different sequences of learning to different children, tests which are designed to measure these learnings must be used as criteria for the success or failure of these programs. The IOX is building a set of objectives, together with criteria for measuring these objectives.

With this kind of data bank, teachers and school administrators can choose those objectives which really describe the program they have found through diagnosing

individual needs of children to be best for their schools. They can use these objectives to develop a program for each child, and test their own program by using the test items developed at the IOX. These tests will provide feedback of a very specific nature to their pupils. Thus, both the student and the teacher will be given information regarding the specific strengths and weaknesses of the pupil--information which can be used for diagnostic and prescriptive purposes. These tests will provide a real learning experience for both the teacher and the student. The student will see precisely those areas in which he has failed to master the material, and the teacher can then devise prescriptive techniques which will permit the student to reach the objectives. The teacher can learn from this kind of information just where in her teaching she has failed to communicate with her children, and where she has achieved the objectives she has set out to attain. By using this kind of a testing procedure, no additional testing is needed. Both the student and the teacher know precisely what has been learned and what additional effort needs to be made in specific areas. This kind of procedure will allow for regular, periodic feedback to the teacher and student throughout the year. This process will preclude the possibility of some child failing to comprehend some objective which keeps him from progressing, since each objective is tested before progressing to the next, and will avoid the feelings of failure which comparative tests tend to develop in children.

Millman gives examples of items which provide the kind of feedback which helps children and teachers learn, and makes maximum use of the test results. The criteria for determining when a child has mastered a particular area needs to be determined empirically as Tyler has suggested (1970). After these tests have been used for a short period of time, this kind of information will be available for the different items and the child will be progressing as fast as he can. The individualizing of instruction, and the non-graded school concepts, which have frequently been hamstrung when tests are given which measure "first grade" or "second grade" when the concept of non-gradedness implies that such entities do not exist, can be avoided. By obtaining test items from the IOX which measure the objectives selected, the child moves according to his own rate. At the end of the year, the teacher can ask for a sampling of items from among all the objectives she has taught during the year to determine the amount of retention. This kind of test would resemble the norm-referenced test but would be specifically designed to meet the needs of an individual school. Cumulations of these results could be maintained at UCLA and data presented much as the National Assessment in Education Project now provides them for the nation, in terms of percentages of children who have successfully accomplished particular objectives. To avoid the problem of recording this data by grades, where schools are on a non-graded program, it would be possible to record the data by age level of the children

being tested on each of the objectives.

This process will also force teachers, parents, school administrators, and the community to decide precisely what the goals of eduction for their children are, and avoid the tendency to let a test determine the curriculum. It will also preclude one problem which faces many school districts, that of setting unrealistic goals for their schools. By specifying exactly what the objectives are that each class in each school is planning to reach during the year, the teacher will be ready to be accountable because she will be setting the goals and criteria against which she will be judged. The community will not expect more than is possible, something which we educators have frequently let them come to expect in the past,because the precision with which the statement of objectives is made will preclude any expectations which are not agreed upon prior to the beginning of the school year.

The use of the IOX will then provide information to both the pupils and teachers which is much more useful to both. The one weakness that might be pointed out is that it will not provide the comparison with other schools. As was mentioned earlier, this disadvantage can be overcome by having the Center for the Study of Evaluation keep careful records of how many children were able to reach each of the objectives during the year for each district and each school which asks for objectives and test items, a process not impossible with today's large computers.

The information which this kind of program will supply to parents is of much greater value to them than the comparative scores which they have been conditioned to expect. By reporting to parents exactly what their children know, parents will tend to reinforce that learning. When specific areas which need additional emphasis have been identified, the parents can be informed and taught to provide the kind of help they can give their children to overcome these weaknesses. This process can help bring the school and community into closer cooperation, also.

Many parents, school administrators, and state officials rightfully want to know something about how their children are doing with reference to other groups, however. I have heard many parents of disadvantaged children tell me that they have been tricked or fooled into believing that their children are progressing well, and then suddenly find out that these same children fail to pass college entrance exam tests or other tests which are used for screening purposes. Parents, especially, should be kept informed fully of the progress of their children, and if the IOX tests are used, some means of comparing how children in individual schools are progressing in comparison with other children of the same age must be found. Two alternatives are offered here.

First, the IOX could provide data on a yearly basis regarding how many children at each age level were able to successfully pass the criteria used to measure the

specific objectives listed in their exchange. This data could be presented in the same form used by the National Assessment in Education Program (NAEP, 1971). That is, percentages of children who perform above or below the mean could be given by schools, school districts, or regions. The data could be used to describe the exact learning level of children of different ages throughout the state, and differences between regions, school districts and schools could be readily seen. The advantage that this kind of data would have over that given by standardized tests is that it tells much more precisely what areas are strengths and what areas are weaknesses in the repertoire of learnings of the children. That is, instead of knowing that children are reading .5 year below grade level, we would be able to say that a certain percentage of children six years old have not mastered the medial sounds in words, but that 90 per cent of those tested at this age level were able to recognize and use the beginning and ending sounds.

A second alternative would be the use of some standardized test which measured as nearly as possible the objectives of most of the school districts within the state on a sampling basis. As mentioned earlier in the paper researchers are all aware that a ten per cent sample of the population will provide information which is within very small limits the same as that which could be gained from testing the total population (Sax, p. 143). Gallup uses only a very small sample of the total population to make his predictions, for example, and the National

Assessment only samples from each of the regions in the United States.

In addition to sampling from the total population, a statewide assessment is not necessary every year. The NAEP and many school districts, such as the Denver Public Schools, test on a three year cycle. If a weakness in the instructional program is discovered by the testing program, it normally takes at least three years before any programanatic change can be implemented which will be reflected in a testing program. I think that this fact is one which researchers should be very careful to clarify as often as possible. Scriven's use of the words formative and summative research have helped to clarify this process. We need to recognize that the first year of any new program is really devoted to a formative type of research--a period of time which is, or should be, devoted to getting the bugs out of a new program. Even the most carefully planned program, when implemented, will have certain constraints or resources which were not originally planned (Provus, 1970). These will affect the program, and it is only when a viable, tested program is ready for dissemination that it should be put into practice. This process then would require a year to develop the program and test it with process evaluations, a second year to implement it widely and de-bug it, and the third year it could reasonably be tested.

If the latter course, sampling the population and testing at three year intervals, were instituted in California, not only would there be a great savings in time

and money, but greater use of the results of the testing could be made. With the masses of data which are available, much time which could more profitably be used to interpret data is used in simply trying to record it, even with the use of high speed computers. A much more important feature in terms of savings, however, is in the time which teachers would save for teaching and instructing children, since the testing time would also be used as part of the teaching-learning process.

A third alternative is to have local school districts develop their own test banks. These test banks would contain items which measure the curriculum which is being used in that district, and would pre-suppose that teachers and administrators in the district had a good understanding of tests and measurements. In fact, I would recommend that no teacher be allowed in the classroom unless he or she had had a course in tests and measurements in which he had demonstrated a proficiency in understanding and interpreting tests. This course should be one in which both the limitations and appropriate uses of both standardized and criterion-referenced tests were fully explored. At present, most texts dealing with tests and measurements fail to include sufficient mention of the criterion-referenced test, which is the most useful tool to the classroom teacher and pupil. It is the understanding of how to appropriately use both the tests which allows the student to get the feedback to enable him to know where he has learned and where he needs to do additional work. Likewise, the teacher knows

he needs to do additional work. Likewise, the teacher knows just where in her curriculum she needs to provide extra help for individual children.

The local districts, having developed the expertise within its teaching staff to prepare criterion referenced tests, should develop a bank of test items for all courses. Lessinger, as long ago as 1963, advocated developing these banks with test items based on Bloom's taxonomy. The new taxonomies in the affective and psycho-motor domain should also be used, although it is recognized that both of these areas are at present in a state of development which makes the writing of test items which are valid and reliable much more difficult.

The advantages of having such a test bank developed at the local level are many: first, the test development skills of the staff will be enhanced. By requiring an in-service course in tests and measurements of all teachers who either have not had one or whose course was totally devoted to interpreting standardized tests, a much more appropriate use can be made of the results of any testing program. Second, if the test items are developed in terms of Bloom's taxonomy in the cognitive area, teachers will have a much better understanding of how to diagnose the areas of difficulty individual children are experiencing. Third, parents will be intimately acquainted with curriculum and the resultant test bank since they will be receiving reports which tell how much their children are achieving in specific areas rather than how they rank with

others in the class. Fourth, the administrators will know precisely what is being taught and learned in the school system, not in terms of broad statements such as "Reading Vocabulary", but in more specific terms such as "90% of the six year old children have a reading vocabulary of 400 words or more".

The disadvantages of this program are obviously that it would be difficult to compare how one school system is doing in comparison with another. This disadvantage could be removed by using the three year cycle of standardized testing along with the locally constructed criterion-referenced tests, or, after a few years, developing a bank of test items which are submitted by local districts following the taxonomy of Bloom and/or Krathwohl to some state agency which would then develop tests based on these items.

These three alternatives are neither mutually exclusive nor exhaustive. This paper has simply posed them as alternatives to the present program to provide a basis which would make others think about the problems and benefits of a state-mandated testing program. In terms of the cost of the present program, and the dire financial straits of many school districts, some thought should be given to a program which is this costly and whose benefits could be achieved at much less cost by other alternatives.

References

Barro, Stephen M. "An Approach to developing Accountability Measures for the Public Schools". Phi Delta Kappan, December, 1970.

Dyer, Henry S. "Statewide Evaluation - What are the Priorities?", Phi Delta Kappan, June 1970.

"Toward Objective Criteria of Professional Accountability in the Schools of New York City", Phi Delta Kappan, December, 1970.

Elliott, Merle, "Testing Programs in Elementary School Districts". California Journal of Educational Research, November, 1956.

Glasser, William. Schools Without Failure, Harper and Row, 1969.

Lessinger, Leon. "Test Building and Test Banks Through the Use of the Taxonomy of Education Objectives". California Journal of Educational Research, November, 1963.

Lyman, Howard B. Test Scores and What They Mean. Prentice-Hall, Englewood Cliffs, New Jersey, 1963.

McAloon, Ann and Ritenour, Louise. Academic Evaluation Program Report X. Educational Testing Service, 1970.

McMurrin, Sterling M. "Society in Predicament, warns Sterling McMurrin." University of Utah Review, March 1971.

Millman, Jason. "Reporting Student Progress: a Case for a Criterion-referenced Marking System, Phi Delta Kappan, December, 1970.

Moellenberg, Wayne P. "National Assessment: Are We Ready?", The Clearing House, April, 1969.

National Assessment of Educational Progress. Report No. 5, Education Commission of the States, Denver, Colorado, 1971.

Provus, Malcolm. "Toward a State System of Evaluation" Journal of Research and Development in Education, Summer, 1970.

Sax, Gilbert. Empirical Foundations of Educational Research, Prentice-Hall, Englewood Cliffs, New Jersey, 1969.

Stake, Robert E. and Wardrop, James L. Gain Scores in Performance Contracting, University of Illinois, Urbana-Champaign, 1971.

Storey, Arthur G. The Measurement of Classroom Learning. Science Research Associates, 1970.

Taylor, Edward A. and Crandall, James H. "A Study of Norm-Equivalence of Certain Tests Approved for the California State Testing Program". California Journal of Educational Research, September, 1962.